

10/581609

JAP3 Rec'd PCT/PTO 05 JUN 2006

## DESCRIPTION

Method of and apparatus for coding moving picture, and  
method of and apparatus for decoding moving picture

Technical field:

5 The present invention relates to a method of and an apparatus for  
coding a moving picture, and a method of and an apparatus for decoding a  
coded moving picture.

Background art:

Various processes for coding moving pictures are known. Of these  
10 processes, a subband coding process is a process for frequency-dividing a  
picture signal and coding signals in respective frequency bands, i.e.,  
subband signals. Unlike block-base orthogonal transform such as discrete  
cosine transform, the subband coding process has features that it does not  
produce block distortions in principle and divides a low-frequency component  
15 recursively for easy hierarchical coding. In the field of still images, JPEG  
2000, which is an international standard for coding process, employs a  
subband coding process using wavelet transform.

If a subband coding process is applied to code a moving picture, then it  
is necessary to take into account not only spatial correlation of signals, but  
20 also temporal correlation of signals. For subband moving picture coding,  
there have mainly been proposed two processes, i.e., a process for  
performing motion compensation on an original picture in spatial domain to  
remove a temporal correlation and thereafter performing subband coding on  
each frame (see, for example, J. R. Ohm, "Three-dimensional subband  
25 coding with motion compensation," IEEE Trans. Image Processing, Vol. 3,  
pp. 559-571, Sept. 1999), and a process for performing subband dividing on

an original picture and thereafter performing motion compensation on each subband region to remove a temporal correlation.

Fig. 1 is a flowchart illustrative of a conventional coding process with motion compensation in spatial domain. A coding process for a set  $A^{(0)}[i]$

5 (0 ≤ i < n, n is the power of 2) of successive frames will be described with reference to Fig. 1.

In steps 201, 202, j=0, i=0, 2, ..., n-2. In steps 203 to 205, successive two frames  $A^{(0)}[i]$ ,  $A^{(0)}[i+1]$  are temporally divided into subbands, producing  $A^{(1)}[i]$  in the low-frequency band and  $E[i+1]$  in the high-frequency band. Then,

10 in step 206, 1 is added to j, and successive signals  $A^{(1)}[i<<1]$ ,  $A^{(1)}[(i+1)<<1]$  in the low-frequency band are temporally divided into subbands, producing  $A^{(2)}[j<<1]$  in the low-frequency band and  $E[(i+1)<<1]$  in the high-frequency band. The above sequence is repeated until frames other than the first frame are coded as a signal in the high-frequency band, i.e., until (1 < j) 15 becomes n, as indicated in step 207. Thereafter, in step 208,  $A^{(0)}[0]$ ,  $E[i]$  (0 < j < n) are spatially divided into subbands and coded.

In the temporal subband division between two frames, a signal in the high-frequency band corresponds to an error signal for motion compensation prediction and a signal in the low-frequency band to an average signal of

20 motion-compensated two frames. Upon a decoding processing, the above process is reversed to spatially combine subband signals for each frame and thereafter temporally combine subbands according to the frame reference relationship. In three-dimensional wavelet coding, frame signals produced by partially combining subbands without using subbands of high-frequency 25 components can be temporally combined in subbands to obtain a decoded picture at a reduced resolution. In this case, since the corresponding

relationship between pixels in frames in motion compensation has to be maintained, the motion information obtained at an original resolution is used as it is except that it is reduced only in scale.

Disclosure of the invention:

5 Problems to be solved by the invention:

According to scalable coding, a stream having a low bit rate can be generated from an original stream by removing codes corresponding to subbands of high-frequency components from the original stream. When the newly generated stream is decoded, a picture represented by the reduced 10 input signal is reconstructed. According to the conventional technology, the motion information obtained at an original resolution is used for decoding at a reduced resolution. Therefore, the amount of codes required for motion information is relatively increased, resulting in a reduction in coding efficiency. Especially, since most of the amount of codes is assigned to motion 15 information at a low bit rate, the picture quality is made lower than if scalability is not applied.

Conversely, with motion information determined at a reduced resolution so as to be optimized for a low bit rate, the coding efficiency at an original resolution is lowered. If the amount of codes required for motion information 20 is reduced according to a process such as integrating coefficient information as it is, then the quality of pictures is greatly reduced due to motion compensation discrepancies.

Therefore, it is an object of the present invention to provide a moving picture coding technology for achieving a higher coding efficiency in a coded 25 stream having a hierarchical structure than the conventional technology while maintaining the hierarchical structure.

Another object of the present invention is to provide a moving picture coding technology for suppressing a reduction in the quality of pictures due to motion compensation discrepancies.

Still another object of the present invention is to provide a technology 5 for decoding moving picture data which have been coded by such a moving picture coding technology.

Means for solving the problems:

According a first aspect of the present invention, a method of coding a moving picture is a method of coding a moving picture for performing 10 hierarchical coding, and comprises the steps of performing a first process on an input picture signal and thereafter spatially dividing the input picture signal into layers to obtain a first signal, reducing the input picture signal with a resolution converting filter and thereafter performing a second process on the input picture signal at a reduced resolution to obtain a second signal, and 15 coding the first signal and the second signal.

According a second aspect of the present invention, a method of coding a moving picture is a method of coding a moving picture for performing hierarchical coding, and comprises the steps of performing a temporal-spatial hierarchical dividing process to divide an input picture signal 20 into a first signal which is obtained by performing a first process on the input picture signal and thereafter spatially dividing the input picture signal into layers and a second signal which is obtained by performing a second process at a reduced resolution on a reduced input picture signal which is produced when the input picture signal is reduced by a resolution converting 25 filter, and recursively performing the temporal-spatial hierarchical dividing process on the reduced input picture signal and thereafter coding signals in

the respective layers.

In these inventions, for example, the first and second processes comprise first and second temporal filtering, respectively, the first signal comprises a temporally filtered lower-layer signal, and the second signal comprises a higher-layer temporally filtered signal. Alternatively, the first process comprises a first motion compensating process, the second process comprises a second motion compensating process, the first signal comprises a prediction error lower-layer signal, and the second signal comprises a higher-layer prediction error signal.

10 According a third aspect of the present invention, a method of coding a moving picture comprises the step of performing, a plurality of times, a three-dimensional subband dividing process for temporally dividing an input picture signal into subbands and spatially dividing the input picture signal into subbands, the three-dimensional subband dividing process comprising the motion information calculating step of calculating motion information representative of a motion between frames of the input picture signal and between bands of an intraband signal which is a band signal of one of low-frequency subbands produced by dividing the input picture signal into subbands, the temporal subband dividing step of temporally dividing the input picture signal and the intraband signal into subbands after the input picture signal and the intraband signal are motion-compensated according to the motion information obtained in the motion information calculating step, thereby generating a temporal low-frequency subband signal and a temporal high-frequency subband signal, the temporal high-frequency subband signal 20 spatially dividing step of spatially dividing temporal high-frequency subband signal into subbands, thereby generating a temporal high-frequency, spatial

25

low-frequency subband and a temporal high-frequency, spatial high-frequency subband, the temporal low-frequency subband signal spatially dividing step of spatially dividing temporal low-frequency subband signal into subbands, thereby generating a temporal low-frequency, spatial low-

- 5 frequency subband and a temporal low-frequency, spatial high-frequency subband, and the band signal spatially dividing step of spatially dividing the intraband signal into subbands, thereby generating a low-frequency intrasubband and a high-frequency intrasubband, wherein the temporal subband dividing step, the temporal high-frequency subband signal spatially dividing step, the temporal low-frequency subband signal spatially dividing step, and the band signal spatially dividing step are performed on the input picture signal, the low-frequency intrasubband obtained after the band signal spatially dividing step is used as the intraband signal, and the temporal subband dividing step, the temporal high-frequency subband signal spatially dividing step, the temporal low-frequency subband signal spatially dividing step, and the band signal spatially dividing step are recursively repeated, and each time these steps are repeated, the temporal low-frequency, spatial low-frequency subband and the temporal high-frequency, spatial low-frequency subband are replaced respectively with the temporal low-frequency subband
- 10 signal and the temporal high-frequency subband signal that are obtained in the temporal subband dividing step performed immediately thereafter.
- 15
- 20

According another aspect of the present invention, a method of decoding a moving picture is a method of decoding a moving picture to decode hierarchical coded data, and comprises the steps of decoding a first signal processed by a first process, a third signal produced when a second signal produced from a second process is spatially divided into layers, and

processing information representing the second process, generating a fourth signal from the first signal and the processing information, and combining the third signal and the fourth signal with each other and thereafter performing inverse transform of the second process to obtain a decoded picture.

5 In this decoding method, for example, the first and second processes comprise first and second temporal filtering, respectively, the first signal comprises a higher-layer temporally filtered signal, the second signal comprises a temporally filtered signal, the third signal comprises a temporally filtered lower-layer signal, the fourth signal comprises a temporally filtered higher-layer signal, and the processing information comprises temporally filtering information. Alternatively, the first and second processes comprise first and second motion compensating processes, respectively, the first signal comprises a higher-layer prediction error signal, the second signal comprises a prediction error signal, the third signal comprises a prediction error lower-layer signal, the fourth signal comprises a prediction error higher-layer signal, and the processing information comprises motion information.

10

15

According still another aspect of the present invention, a method of decoding a moving picture to obtain a decoded picture by combining layers of hierarchical coded data for each frame and thereafter temporally inverse-filtering the data, comprises the steps of decoding a higher-layer temporally filtered signal which is produced by first temporal filtering, a temporally filtered lower-layer signal produced when a temporally filtered signal which is produced by second temporal filtering is spatially divided into layers, and temporal filtering information representing the second temporal filtering,

20

25

generating a temporally filtered higher-layer signal from the higher-layer temporally filtered signal and the temporal filtering information, performing a

temporally filtered signal combining process to combine the temporally filtered higher-layer signal and the temporally filtered lower-layer signal to generate a combined temporally filtered signal, and producing a decoded picture by regarding the combined temporally filtered signal as the higher-  
5 layer temporally filtered signal, decoding the temporal filtering information and the temporally filtered lower-layer signal in a layer lower than a layer of interest, recursively performing the temporally filtered signal combining process, and thereafter performing temporally inverse-filtering.

According yet another aspect of the present invention, a method of  
10 decoding a moving picture comprises the step of generating a decoded picture signal according to a three-dimensional subband combining process for spatially combining subband signals for each frame and thereafter performing temporal subband combining process for temporally combining a temporal low-frequency subband and a temporal high-frequency subband,  
15 the three-dimensional subband combining process comprising the temporal high-frequency subband combining step of generating a combined temporal high-frequency subband signal by referring to a temporal high-frequency, spatial low-frequency signal which is a spatial low-frequency signal of a temporal high-frequency subband, and a temporal high-frequency, spatial  
20 high-frequency subband which is subband of a high-frequency band adjacent to the low-frequency signal, and additionally both or either one of a temporal low-frequency, spatial low-frequency subband which is in the same frequency band as the temporal high-frequency, spatial low-frequency signal, and a temporal low-frequency, spatial high-frequency subband which is a  
25 subband of a high-frequency band adjacent to the subband signal, and motion information representing a motion compensating process

corresponding to the temporal high-frequency subband, the temporal low-frequency subband spatially combining step of combining the temporal low-frequency, spatial low-frequency subband and the temporal low-frequency, spatial high-frequency subband, and the temporally combining step of

- 5 performing a motion compensation predicting process on the temporal low-frequency subband and the temporal high-frequency subband, and thereafter performing temporal subband combination, wherein the temporal high-frequency subband combining step is performed on the temporal high-frequency, spatial low-frequency signal which is in the lowest frequency band
- 10 of the temporal high-frequency subband, and the temporal low-frequency subband spatially combining step is performed on the temporal low-frequency, spatial low-frequency subband which is in the lowest frequency band of the temporal low-frequency subband, and the band signal obtained by the temporal high-frequency subband combining step is regarded as a
- 15 new temporal high-frequency, spatial low-frequency signal, and the band signal obtained by the temporal low-frequency subband spatially combining step is regarded as a new temporal low-frequency, spatial low-frequency subband, the temporal high-frequency subband spatially combining step and the temporal low-frequency subband spatially combining step are recursively
- 20 repeated, producing the temporal low-frequency subband and the temporal high-frequency subband.

According to the present invention, a coded stream having a hierarchical structure is motion-compensated based on motion information which is different from layer to layer. According to the present invention, for

25 reconstructing low-rate coded data, except for a high-frequency component, of coefficient information, motion information corresponding to a motion

compensation at a high resolution is deleted for achieving a higher coding efficiency than heretofore while maintaining the hierarchical structure. A reduction in the picture quality due to motion compensation discrepancies is greatly reduced by correcting a low-frequency component based on the

5 motion compensation at the high resolution.

Brief description of the drawings:

Fig. 1 is a flowchart of the processing sequence of a conventional coding process with motion compensation in spatial domain;

10 Fig. 2 is a diagram showing an arrangement of a moving picture coding apparatus according to an embodiment of the present invention;

Fig. 3 is a flowchart of the processing sequence of a moving picture coding method according to the embodiment of the present invention;

Fig. 4 is a flowchart of a process of temporal-spatial subband division of two frames in Fig. 3;

15 Fig. 5 is a diagram showing an arrangement of a moving picture decoding apparatus according to an embodiment of the present invention;

Fig. 6 is a flowchart of the processing sequence of a moving picture decoding method according to the embodiment of the present invention;

Fig. 7 is a flowchart of a process of temporal-spatial subband combination of two frames in Fig. 6; and

20 Fig. 8 is a schematic diagram showing a computer system for implementing the arrangements of the moving picture coding apparatus and the moving picture decoding apparatus.

Best mode for carrying out the invention:

25 According to the present invention, for hierarchical coding, a temporal filtering lower-layer signal obtained by performing first temporal filtering and

thereafter spatial hierarchical division, and a higher-layer temporal filtering signal obtained by reducing an input picture signal with a resolution converting filter and thereafter performing second temporal filtering at a reduced resolution, are coded. Alternatively, according to the present

5 invention, a prediction error lower-layer signal obtained by performing first motion compensation and thereafter spatial hierarchical division, and a higher-layer error prediction signal obtained by reducing an input picture signal with a resolution converting filter and thereafter performing second temporal filtering at a reduced resolution, are coded. Namely, the present

10 invention is characterized in that motion compensation is effected on a coding stream having a hierarchical structure based on motion information that is different between the layers. Here, motion information refers to information with respect to a translation of each of blocks of fixed size or variable size which make up a frame, or information with respect to a

15 geometrical transformation such as an affine transformation into each of small areas making up a frame, or information with respect to a geometrical transformation such as an affine transformation on a frame in its entirety.

Specific embodiments of the present invention will be described below.

First, moving picture coding according to the present invention will be

20 described below.

Fig. 2 shows an arrangement of a moving picture coding apparatus according to an embodiment of the present invention, and also shows a hierarchical structure of a signal produced after two frames, which are part of an input picture signal, are temporally and spatially divided into subbands.

25 The moving picture coding apparatus comprises first to third dividers 1001 to 1003 for performing temporal subband division, fourth and fifth

dividers 1004, 1005 for performing spatial subband division, and first and second low-pass filters 1006, 1007. Input picture signal 10 is supplied to first divider 1001 and first low-pass filter 1006. Output 11 of first divider 1001 is supplied to fourth divider 1004. Output 20 of first low-pass filter 1006 is 5 supplied to second divider 1002 and second low-pass filter 1007. Output 21 of second divider 1002 is supplied to fifth divider 1005. Output 30 of second low-pass filter 1007 is supplied to third divider 1003.

In the moving picture coding apparatus, input picture signal 10 is temporally divided into subbands by first divider 1001, and thereafter 10 spatially divided into subbands at a single stage, thereby generating low-frequency subband signal 12 and high-frequency subband signal 13. When input picture signal 10 passes through first low-pass filter 1006, intrasubband signal 20 is generated. Intrasubband signal 20 is temporally divided into subbands by third divider 1002, generating low-frequency temporal subband 15 signal 21. Low-frequency subband signal 12 generated by fourth divider 1004 is replaced with low-frequency temporal subband signal 21. Namely, the results of the single-stage hierarchical division of input picture signal 10 are high-frequency subband signal 13 and low-frequency temporal subband signal 21 according to the . Similarly, the results of the single-stage 20 hierarchical division of low-frequency temporal subband signal 21 are high-frequency subband signal 23 of low-frequency temporal subband signal 21 and low-frequency temporal subband signal 31 which is generated by temporally dividing low-frequency subband signal 30 of intrasubband signal 20. The above hierarchical division is performed recursively to realize a 25 multiple hierarchical structure.

Each of low-pass filters 1006, 1007 may comprise either one of a

general down-sampling filter for reducing a resolution horizontally and vertically to 1/2 and a low-pass filter in fourth and fifth dividers 1004, 1005 which perform spatial subband division. Hereinafter, a coding process with such a hierarchical structure will be described below on the assumption that

5 the low-pass filters for spatial subband division are used.

Fig. 3 is a flowchart showing a coding process according to the present embodiment. The coding process for coding a set  $A^{(0)}[i]$  ( $0 \leq i < n$ ,  $n$  is the power of 2) of successive frames used as an original picture input will be described below with reference to Fig. 3.

10 First, in steps 101, 102,  $j = 0, 1, 2, \dots, n-2$ . In step 103, successive two frames  $A^{(0)}[i], A^{(0)}[i+1]$  are temporally and spatially divided into subbands, producing subband signals  $A^{(1)}[i], E^{(1)}[i+1]$  and motion information  $V[i+1]$ .

15 Fig. 4 is a flowchart of the processing sequence of the temporal-spatial subband division of two frames in step 103 shown in Fig. 3. It is assumed that frame  $B0$  is a past frame with respect to frame  $C0$ , and a general process of temporal-spatial subband division on frames  $B0, C0$  will be described below with reference to Fig. 4.

20 First, in step 111, a motion of frame  $B0$  with respect to frame  $C0$  is estimated to produce motion information  $V0$ . A motion refers to a translation of each of blocks of fixed size or variable size which make up the frame, or a geometrical transformation such as an affine transformation into each of small areas making up a frame, or a geometrical transformation such as an affine transformation on a frame in its entirety.

25 Next, in step 112, based on motion information  $V0$ , frames  $B0, C0$  are temporally divided into subbands to generate low-frequency subband  $A0$  and high-frequency subband  $E0$ . As a temporal subband division process, a

process disclosed in A. Secker et al., "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," IEEE Trans. Int. Conf. Image Proc., pp. 1029-1032, October, 2001 will be described below.

5 If it is assumed that a pixel value at intraframe coordinates [p,q] in frame B0 is represented by B0[p,q], a pixel value at intraframe coordinates [p,q] after frame B0 has been motion-compensated based on the motion estimated in step 111 by W<sub>B0</sub>(B0)[p,q], and a pixel value at intraframe coordinates [p,q] after frame C0 has been motion-compensated by

10 W<sub>C0</sub>(C0)[p,q], then the following equations are satisfied:

$$E0^*[p,q] = 1/2(C0[p,q] - W_{B0}(B0)[p,q]) \quad (1)$$

$$A0^*[p,q] = B0[p,q] + W_{C0}(E0^*)[p,q] \quad (2)$$

According to another temporal subband division process, if a filter having temporal filter length longer than 2 is used, then assuming that filters for dividing a plurality of input frames B0<sub>i</sub> into low- and high-frequency bands are represented respectively by f<sub>l</sub>[i] (0 ≤ i < n<sub>l</sub>), f<sub>h</sub>[i] (0 ≤ i < n<sub>h</sub>), A0<sup>\*</sup> and E0<sup>\*</sup> are expressed as follows:

$$A0^*[p,q] = \sum_{i=0}^{n_l-1} f_l[i] \cdot W_{B0i}(B0_i)[p,q] \quad (1')$$

$$E0^*[p,q] = \sum_{j=0}^{n_h-1} f_h[j] \cdot W_{B0j}(B0_j)[p,q] \quad (2')$$

20 L. Lio et al., "Motion Compensated Lifting Wavelet And Its Application in Video Coding," IEEE Int. Conf. Multimedia & Expo 2001, Aug., 2001 shows motion compensation upon the processing of each filter in a lifting process for realizing high-order subband division with a superposition of primary filters. According to the disclosed process, if even-numbered frames

of a plurality of input frames are represented by  $B_0$ ; and odd-numbered frames by  $C_0$ ; then frames  $B'_0$ ,  $C'_0$  after being multiplied by primary filters are expressed with constants  $\alpha$ ,  $\beta$  as follows:

$$C'_0[p,q] = C_0[p,q] + \alpha (W_{B0i}(B_0[p,q] + W_{B0i+1}(B_0[p,q]))[p,q]) \quad (1'')$$

5       $B'_0[p,q] = B_0[p,q] + \beta (W_{C0i}(C_0[p,q] + W_{C0i+1}(C_0[p,q]))[p,q]) \quad (2'')$

By alternately repeating the two filter processes, temporal subband division using the lifting process is performed. There is known another process which is equivalent to the ordinary motion compensation prediction without generating  $A_0$  of low-frequency components.

10      After  $A_0$ ,  $E_0$  are obtained, they are spatially divided into subbands once in step 113.

If dual frequency division is performed as subband division using a one-dimensional filter bank, then there are generated four subbands, i.e., a subband divided both horizontally and vertically into low-frequency bands, a subband divided horizontally into a low-frequency band and vertically into a high-frequency band, a subband divided horizontally into a high-frequency band and vertically into a low-frequency band, and a subband divided both horizontally and vertically into high-frequency bands. These subband transforms are defined respectively as  $LL()$ ,  $LH()$ ,  $HL()$ ,  $HH()$ . A set of three subbands  $LH(C_0)$ ,  $HL(C_0)$ ,  $HH(C_0)$  is defined as  $H(C_0)$ . According to these definitions,  $LL(A_0)$ ,  $H(A_0)$ ,  $LL(E_0)$ ,  $H(E_0)$  are obtained.

Thereafter, in step 115, frames  $B_0$ ,  $C_0$  are spatially divided into subbands in one layer, producing  $LL(B_0)$ ,  $H(B_0)$ ,  $LL(C_0)$ ,  $H(C_0)$ .  $LL(B_0)$ ,  $LL(C_0)$  are defined as  $B_1$ ,  $C_1$ , respectively. In step 116, motion information  $V_1$  representing a motion compensation between these subbands is newly calculated.

Motion information calculating processes include a process for estimating a motion anew and a process for integrating some motion information corresponding to B0, C0. Particularly, hierarchical coding with coefficient code information and motion information being associated with each other can be realized by performing hierarchical coding on motion information and using motion information corresponding to only its base layer.

According to a process for performing hierarchical coding on V1, V2, for example, motion information obtained by estimating a motion in a picture having a reduced resolution is represented by V2, motion information obtained by estimating a motion in an original picture by V1, and information produced by subtracting a twofold of V2 from V1 and V2 are coded.

Furthermore, as with subband coding, motion information is divided into subbands in x and y directions of the picture to provide a hierarchical representation of the motion information.

In step 117, based on the information thus obtained, B1, C1 are temporally divided into subbands, producing low-frequency subband A1' and high-frequency subband E1'. It should be noted that A1' is not equal to LL(A0') and E1' is not equal to LL(E0').

After A1', E1' are obtained, if the number of spatial subband divisions is 1 in step 118, then A1' is used as the division result instead of LL(A0'), H(B0) is used as the divided result instead of H(A0'), and E1' is used as the divided result instead of LL(E0'). If the number of spatial subband divisions is not 1, then A1', E1' are spatially divided into subbands once, generating LL(A1'), H(A1'), LL(E1'), H(E1') in step 119. Thereafter, control goes back to step 115 in which B1, C1 are divided into subbands once. In step 116, motion information V2 is calculated with respect to obtained B2, C2.

Thereafter, temporal subband division with motion compensation is performed in step 117.

The above process is carried out until the number of divisions becomes m as shown in step 118. Then, in step 121, obtained  $LL(A_m)$ ,  $H(A_k)$ ,

5  $LL(E_m)$ ,  $H(E_k)$  ( $0 \leq k < m$ ) are used as the division results. In step 122,  $V_k$  ( $0 \leq k < m$ ) is output as motion information of the entire subband division on the two frames, after which the process is put to an end. In this manner, the subband division in step 103 is finished.

10 The coding process in the present embodiment shown in Fig. 3 will further be described below.

After step 103,  $A^{(0)}[0]$ , which is temporal low-frequency subband, is subjected to spatial subband combination to generate  $A^{(1)}[0]$  in step 105. This is to allow  $A^{(1)}[0]$  to be temporally divided into subbands according to the processing in step 103 in a temporal layer that is one level higher.

15 In steps 106, 107, the processing in steps 103, 105 is performed on  $A^{(0)}[n-2]$ ,  $A^{(0)}[n-1]$ . Thereafter, in step 108, j is incremented by 1. With  $i = 0$ ,  $2, \dots, n/2-2$ , the temporal subband division of  $A^{(1)}[i<<1]$  and  $A^{(1)}[(i+1)<<1]$  (step 103) and the spatial subband combination of  $A^{(1)}[i<<1]$  (step 105) are repeated.

20 The above processing loop is performed until j becomes equal to  $\log 2(n)-1$ .

If j, which represents the present number of temporal divisions, is equal to  $\log 2(n)-1$  when step 103 is ended, i.e., in step 104, then the temporal-spatial subband division of all signals is finished. According to the coding process, in step 109, obtained signals  $A^{(0)}[0]$ ,  $E^{(i)}$  ( $0 < i < n$ ) are quantized and losslessly coded. Linear quantization, nonlinear quantization, and vector

quantization are used as the quantization process, and in addition to these processes, bit-plane quantization used in JPEG 2000 which is an international standard for still image coding may be also used. Zero-tree coding disclosed in J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients", IEEE Trans. Signal Processing, vol. 41, pp. 3445-3462, Dec. 1993, arithmetic coding, or run-length coding may be used as the lossless coding. In step 110,  $V[i]$  ( $0 \leq i < n$ ) are coded. The coding process for  $A^{(0)}[k]$  ( $0 \leq k < n$ ) is now put to an end.

According to the moving picture coding process described above, a 10 three-dimensional subband dividing process for temporally dividing an input picture signal into subbands and spatially dividing the input picture signal into subbands is performed a plurality of times. The subband dividing process comprises:

- the motion information calculating step of calculating motion 15 information representative of a motion between frames of an input picture signal and between bands of an intraband signal which is a band signal of one of low-frequency subbands produced by dividing the input picture signal into subbands;
- the temporal subband dividing step of temporally dividing the input 20 picture signal and the intraband signal into subbands after the input picture signal and the intraband signal are motion-compensated according to the motion information obtained in the motion information calculating step, thereby generating a temporal low-frequency subband signal and a temporal high-frequency subband signal;
- the temporal high-frequency subband signal spatially dividing step of 25 spatially dividing the temporal high-frequency subband signal into subbands,

thereby generating a temporal high-frequency, spatial low-frequency subband and a temporal high-frequency, spatial high-frequency subband;

the temporal low-frequency subband signal spatially dividing step of spatially dividing temporal low-frequency subband signal into subbands,

- 5 thereby generating a temporal low-frequency, spatial low-frequency subband and a temporal low-frequency, spatial high-frequency subband; and

the band signal spatially dividing step of spatially dividing the intraband signal into subbands, thereby generating a low-frequency intrasubband and a high-frequency intrasubband.

- 10 The temporal subband dividing step, the temporal high-frequency subband signal spatially dividing step, the temporal low-frequency subband signal spatially dividing step, and the band signal spatially dividing step are performed on the input picture signal. The low-frequency intrasubband obtained after the band signal spatially dividing step is used as the intraband signal, and the temporal subband dividing step, the temporal high-frequency subband signal spatially dividing step, the temporal low-frequency subband signal spatially dividing step, and the band signal spatially dividing step are recursively repeated. Each time these steps are repeated, the temporal low-frequency, spatial low-frequency subband and the temporal high-frequency, spatial low-frequency subband are replaced respectively with the temporal low-frequency subband signal and the temporal high-frequency subband signal that are obtained in the temporal subband dividing step performed immediately thereafter.
- 15
- 20
- 25

According to the present embodiment, the process sequence is a

sequence in which a frame in a certain layer is temporally and spatially divided into subbands, and thereafter the frame to be coded in a next layer is

once subjected to spatial subband combination. However, these two processes can be integrated with each other. The feature of the present invention resides in that a motion compensation is appropriately corrected depending on a spatial frequency band, and the order of spatial subband dividing processes has nothing to do with the objects of the present invention.

5 Moving picture decoding according to the present invention will now be described below. According to the present embodiment, a decoded picture has a resolution that is represented by 1/(the power of 2) of the resolution of an original picture in both temporal and spatial directions. Specifically, if the 10 number of spatial subband divisions in the coding process is represented by  $m$ , then it is possible to reconstruct decoded pictures having horizontal and vertical resolutions represented by  $1/2, 1/4, \dots, 1/2^m$  of the resolution of the original picture. If the number of temporal subband divisions is  $n_0 = \log_2(n)$ , then it is possible to reconstruct decoded pictures having frame rates 15 represented by  $1/2, 1/4, \dots, 1/2^{n_0}$  of the frame rate of the original picture.

Fig. 5 shows an arrangement of a moving picture decoding apparatus according to an embodiment of the present invention. The moving picture decoding apparatus serves to decode moving picture data that have been divided into subbands and coded. Fig. 5 also shows the concept of a 20 subband combination which corresponds to the subband division shown in Fig. 2. Specifically, the moving picture decoding apparatus shown in Fig. 5 is connected to the moving picture coding apparatus shown in Fig. 2 through a signal transmission medium, and receives and decodes moving picture data coded by the moving picture coding apparatus.

25 The moving picture decoding apparatus comprises first to third combiners 2001 to 2003 for performing temporal subband combination,

fourth and fifth combiners 2004, 2005 for performing spatial subband combination, and first and second dividers 2006, 2007 for forming temporal subband division. Third combiner 2003 is supplied with low-frequency temporal subband signal 31 and outputs decoded picture 36. Second divider 5 2007 is supplied with decoded picture 36 and generates signal 24. Fifth combiner 2005 is supplied with signal 24 and high-frequency subband signal 23, and signal 25 which is output from fifth combiner 2005 is supplied to second combiner 2002. Second combiner 2002 outputs decoded picture 26 which is supplied to first divider 2006. First divider 2006 outputs low-10 frequency subband estimated signal 14. Fourth combiner 2004 is supplied with low-frequency subband estimated signal 14 and high-frequency subband signal 13 and outputs signal 15. First combiner 2001 is supplied with signal 15 and outputs decoded picture 16.

With the moving picture decoding apparatus, in order to obtain 15 decoded picture 36 having the smallest reduced resolution, third combiner 2003 may perform temporal subband combination on low-frequency temporal subband signal 31 which is in the lowest band of coded subband signals. In order to obtain decoded picture 26 in a layer lower than decoded picture 36, i.e., decoded picture 26 having a resolution that is one level 20 higher than decoded picture 36, a low-frequency subband corresponding to high-frequency subband signal 23 belonging to that layer is required. Therefore, signal 24 that is produced when decoded picture 36 is temporally divided into subbands by second divider 2007 is used as a low-frequency subband estimated signal. After fifth combiner 2005 performs spatial 25 subband combination of low-frequency subband estimated signal 24 and high-frequency subband signal 23, second combiner 2002 performs temporal

subband combination to produce decoded picture 26. The temporal subband division in second divider 2007 is uniquely determined by the temporal subband combination in second combiner 2002. Similarly, in order to obtain decoded picture 16 having a resolution that is one level higher than 5 decoded picture 26, low-frequency subband estimated signal 14 which is produced from decoded picture and temporal subband division 2006 and high-frequency subband signal 13 may be spatially combined by fourth combiner 2004, after which first combiner 2001 may perform temporal subband combination. Decoded pictures having different resolutions can be 10 obtained by repeatedly performing the above decoding process on subband signals having a hierarchical structure.

The decoding process will be described below with reference to Fig. 6 which shows a flowchart representing details of the decoding process. A processing sequence for reconstructing decoded picture  $A^{(j_0)}_{(k_0)[i]}$  having a 15 horizontal and vertical resolution of  $1/2^{k_0}$  ( $0 \leq k_0 \leq m$ ) and a frame rate of  $1/2^{j_0}$  ( $0 \leq j_0 \leq n_0$ ) from original picture  $A^{(0)}[i]$  ( $0 \leq i < n$ ,  $n$  is the power of 2) will be described below.

First, in step 151,  $j$  is set to  $\log 2(n)-1$ . Thereafter, in step 152, the coded data is subjected to inverse-transform of the lossless coding and 20 inverse-quantization. The resultant signal obtained from this process is defined as  $A^{(n_0)}[0]$ ,  $E^*[i]$  ( $0 < i < n$ ) according to the symbols used in Fig. 3. Then, in step 153, it is determined whether  $j_0$  is equal to  $n_0$  or not. If  $j_0$  is 25 equal to  $n_0$ , then it is not necessary to perform temporal subband combination, and  $A^{(0)}[0]$  is subjected to spatial subband combination in  $k_0$  layers. Thereafter, when decoded picture  $A^{(0)}_{(k_0)[0]}$  has been reconstructed, the decoding process is put to an end. If  $j_0$  is not equal to  $n_0$  in step 153,

then  $A^{(0)}[0]$ ,  $E[n/2]$  are subjected to temporal and spatial combination in steps 155, 156.

Fig. 7 shows the processing sequence for temporally and spatially combining the subbands of two frame data in step 156. It is assumed that 5 the number of subband combinations upon the decoding is represented by  $k_0$ . If  $k_0$  is zero, then the decoded picture is of the same resolution as the original picture. If  $k_0$  is positive, then the decoded picture is of a resolution which has been reduced by the  $k_0$ -th power of 2. The data of two frames to be processed for subband combination have a hierarchical structure that is 10 spatially divided into subbands  $m$  times.

According to step 121 shown in Fig. 4, a subband signal belonging to the lowest frequency band according to the spatial subband division, of subband signals belonging to the low-frequency band according to the temporal subband division, corresponds to  $A_m$ , and a high-frequency 15 subband after the subband division in the  $k$ -th layer corresponds to  $H(A_k)$  ( $0 \leq k < m$ ). Similarly, signals according to the spatial subband division, of subband signals belonging to the high-frequency band according to the temporal subband division can be associated with  $E_m$  and  $H(E_k)$  ( $0 \leq k < m$ ), respectively. Motion information representing a motion compensation for two 20 frames can be associated with  $V_k$  ( $0 \leq k < m$ ) that is output in steps 111, 116 shown in Fig. 4. Motion information may be independently coded or may be hierarchically coded. Therefore, a processing sequence for reconstructing decoded pictures  $B_{k0}$ ,  $C_{k0}$  having a resolution that is reduced by the  $k_0$ -th 25 power of 2 from the original pictures  $B_0$ ,  $C_0$  by referring to  $A_m$ ,  $H(B_k)$ ,  $E_m$ ,  $(E_k)$  ( $0 \leq k < m$ ) will be described below with reference to Fig. 7.

First, in step 171,  $k$  is set to  $k = m$ . In step 172, then  $A_m$ ,  $E_m$  are

subjected to temporal subband combination, thereby producing  $B_m$ ,  $C_m$ .

$$B_m[p,q] = A_m[p,q] - W_{Cm}(E_m)[p,q] \quad (3)$$

$$C_m[p,q] = 2 \times E_m[p,q] + W_{Bm}(B_m)[p,q] \quad (4)$$

where  $W_{Bm}$ ,  $W_{Cm}$  are a filter representing a motion compensation from  $B_m$

5 to  $C_m$  and a filter representing a motion compensation from  $C_m$  to  $B_m$ , and  
are determined by the motion information  $V_m$  used in the coding process  
and an interpolating process.

In step 173, if  $k_0$  is equal to  $m$ , the decoding process is terminated. If  
10  $k_0$  is not equal to  $m$ , then it is necessary to obtain  $LL(A_{m-1})$ ,  $LL(E_{m-1})$  in  
order to perform one subband combination. Therefore, in step 174, using  
motion information  $V_{m-1}$  which is used in a motion compensation for  $B_{m-1}$ ,  
15  $C_{m-1}$  in the  $(m-1)$ -th layer, estimated values  $LL^{est}(A_{m-1})$ ,  $LL^{est}(E_{m-1})$  for  
 $LL(A_{m-1})$ ,  $LL(E_{m-1})$  are calculated.

$$LL^{est}(E_{m-1}) = 1/2(C_m[p,q] - W_{B_{m-1}}^L(B_m)[p,q]) \quad (5)$$

$$15 LL^{est}(A_{m-1}) = B_m[p,q] - W_{C_{m-1}}^L(LL^{est}(E_{m-1}))[p,q] \quad (6)$$

where  $W_{B_{m-1}}^L$ ,  $W_{C_{m-1}}^L$  are motion compensation filters obtained by reducing  
the motion information  $V_{m-1}$  both horizontally and vertically to 1/2 and  
reducing the size of blocks, which are a unit of motion compensation, to 1/2.  
Alternatively, those which are the same as with the hierarchical motion

20 compensation disclosed in T. Kimoto, "Multi-Resolution MCTF for 3D  
Wavelet Transformation in Highly Scalable Video," ISO/IEC  
JTC1/SC29/WG11, M9770, Trondheim, July 2003 are employed.

Specifically, prediction signal  $W_{B_{m-1}}(B_{m-1})$  obtained by a motion  
compensation on  $C_{m-1}$  in the  $(m-1)$ -th layer is represented as the sum of a  
25 signal due to only spatial low-frequency subband  $B_m$  and a signal due to  
only high-frequency subband  $H(B_{m-1})$ . The former is used as  $W_{B_{m-1}}^L(B_m)$

for estimating  $LL(Em-1)$ .

Thereafter, in step 175,  $LL^{est}(Am-1)$ ,  $H(Am-1)$  are subjected to subband combination, and  $LL^{est}(Em-1)$ ,  $H(Em-1)$  are subjected subband combination, thereby producing  $Am-1$ ,  $Em-1$ . As indicated in steps 173,

5 176, the processing from steps 172 to 175 is repeated to obtain subbands  $Bk0$ ,  $Ck0$  corresponding to layer  $k0$ . Then, the temporal-spatial subband combination in step 156 shown in Fig. 6 is terminated.

In the present embodiment, the correction of subbands (step 174) and the spatial subband combination (step 175) are described as independent 10 steps. However, these steps may be integrated by using a filter that is produced by combining the motion compensation filter for subband correction and the subband combining filter. In the present embodiment, temporal subband combination is preformed on  $Ek$ ,  $Ak$  according to the motion information  $Vk$  to obtain  $Bk$ ,  $Ck$ , after which  $LL^{est}(Ek-1)$ ,  $LL^{est}(Ak-1)$  15 are calculated by referring to the motion information  $Vk-1$ . However, if  $Bk$ ,  $Ck$  do not need to be output, then these processes may be integrated with each other.

According to a process which is the same as the process disclosed in T. Kimoto, "Multi-Resolution MCTF for 3D Wavelet Transformation in Highly Scalable Video," ISO/IEC JTC1/SC29/WG11, M9770, Trondheim, July 2003, 20 it is possible to add a process for correcting  $LL^{est}(Ek-1)$  so as to become closer to  $LL(Ek-1)$  by referring to  $H(Ek-1)$ ,  $Ck$ , and correcting  $LL^{est}(Ak-1)$  so as to become closer to  $LL(Ak-1)$  by referring to  $H(Ak-1)$ ,  $Ak$ .

The decoding process will further be described below with reference to 25 Fig. 6. After performing subband combination on  $A^{(0)}[0]$ ,  $E^{(0)}[n/2]$ , pictures  $A^{(0)}_{(k0)}[0]$ ,  $A^{(0)}_{(k0)}[n/2]$  whose resolution is  $1/2^{k0}$  of the resolution of the original

picture are obtained. If  $j_0$  is equal to  $n_0-1$  in step 157, then the decoding process is put to an end here. If  $j_0$  is not equal to  $n_0-1$  in step 157, then for temporal-spatial subband combination in the next repetitive loop, pictures  $A^{(0)}_{(k_0)}[0], A^{(0)}_{(k_0)}[n/2]$  are spatially divided into subbands, generating  $A^{(0-1)}_{(k_0)}[0], A^{(0-1)}_{(k_0)}[n/2]$ .

5     In step 162,  $j$  is decremented by 1. Thereafter, temporal-spatial subband combination in the next layer is performed on  $A^{(0)}[0], E^{(0)}[n/4]$  and  $A^{(0)}[n/2], E^{(0)}[3n/4]$  as indicated in steps 156, 159, 160. The subband combination is repeated in the manner described above, and the decoding process is put to an end when  $j$  becomes equal to 0 in step 161.

10    The above decoding process has the step of generating a decoded picture signal according to a three-dimensional subband combining process for performing spatial subband combination on subband signals for each frame and thereafter performing temporal subband combination on a temporal low-frequency subband and a temporal high-frequency subband.

15    The three-dimensional subband combining process comprises:

      the temporal high-frequency subband combining step of generating a combined temporal high-frequency subband signal by referring to a temporal high-frequency, spatial low-frequency signal which is a signal of a spatial low-frequency band of a temporal high-frequency subband, and a temporal

20    high-frequency, spatial high-frequency subband which is a subband of a high-frequency band adjacent to the low-frequency signal, and additionally both or either one of a temporal low-frequency, spatial low-frequency subband which is in the same frequency band as the temporal high-frequency, spatial low-frequency signal, and a temporal low-frequency,

25    spatial high-frequency subband which is a subband of a high-frequency band adjacent to the subband signal, and motion information representing a

motion compensating process corresponding to the temporal high-frequency subband;

the temporal low-frequency subband spatially combining step of combining the temporal low-frequency, spatial low-frequency subband and

5 the temporal low-frequency, spatial high-frequency subband; and

the temporally combining step of performing temporal subband combination of the temporal low-frequency subband and the temporal high-frequency subband after the temporal low-frequency subband and the temporal high-frequency subband are processed for a motion compensation prediction.

The temporal high-frequency subband combining step is performed on the temporal high-frequency, spatial low-frequency signal which is in the lowest frequency band of the temporal high-frequency subband, and the temporal low-frequency subband spatially combining step is performed on

15 the temporal low-frequency, spatial low-frequency subband which is in the lowest frequency band of the temporal low-frequency subband. The band signal obtained by the temporal high-frequency subband combining step is regarded as a new temporal high-frequency, spatial low-frequency signal, and the band signal obtained by the temporal low-frequency subband

20 spatially combining step is regarded as a new temporal low-frequency, spatial low-frequency subband. The temporal high-frequency subband spatially combining step and the temporal low-frequency subband spatially combining step are recursively repeated. As a result, the temporal low-frequency subband and the temporal high-frequency subband are obtained.

25 In the present embodiment, described is the case in which the frame reference relationship in the temporal subband division is of a hierarchical

structure. However, the present invention is also applicable where the frame reference relationship is of any desired structure.

The present embodiment has been described with respect to a limited arrangement in which a past frame is converted into low-frequency subbands 5 in one temporal subband division. However, the present invention is also applicable where a future frame is converted into low-frequency subbands or two frames are temporally divided as they are predicted bidirectionally. At any rate, low-frequency subbands produced when each of temporally divided subbands is spatially divided are replaced with subbands produced when 10 low-frequency subbands produced from a spatially divided picture to be coded are temporally divided, and correction is made so that the desired decoded results can be obtained using decoded results of frames which are paired when they are decoded or using the subbands.

In the present embodiment, the subband division is employed as a 15 conversion process for realizing hierarchical coding. However, the present invention is applicable to any hierarchical coding processes. In the subband division, a signal corresponding to a low-frequency band is associated with a higher layer. According to the coding process based on the present invention, after an input picture signal is divided into layers, a higher-layer 20 signal produced when a prediction error signal obtained subsequently to an interframe prediction process is divided into layers may be replaced with a prediction error obtained when the higher-layer signal is processed by the interframe prediction process. In the decoding process, a higher-layer of hierarchical frame signals is corrected into a higher-layer signal produced 25 when a prediction error signal obtained from the interframe prediction process performed on the input picture signal is divided into layers.

If the prediction error signal is employed, a three-dimensional subband dividing process in a moving picture coding process comprises:

- the motion information calculating step of calculating motion information representative of a motion between frames of an input picture signal and between bands of an intraband signal which is a band signal of one of low-frequency subbands produced by dividing the input picture signal into subbands;

- the motion compensation predicting step of obtaining a prediction error signal by performing a motion compensation predicting process on the input picture signal and the intraband signal according to the motion information obtained in the motion information calculating step;

- the prediction error signal spatially dividing step of spatially dividing the prediction error signal into subbands, thereby generating a low-frequency prediction error subband and a high-frequency prediction error subband; and
- the band signal spatially dividing step of spatially dividing the intraband signal into subbands, thereby generating a low-frequency intrasubband and a high-frequency intrasubband.

- The motion information calculating step, the motion compensation predicting step, the prediction error signal spatially dividing step, and the band signal spatially dividing step are performed on the input picture signal. The low-frequency intrasubband obtained after the band signal spatially dividing step is used as the intraband signal, and the motion information calculating step, the motion compensation predicting step, the prediction error signal spatially dividing step, and the band signal spatially dividing step are recursively repeated. Each time these steps are repeated, the low-frequency prediction error subband obtained by the prediction error signal

spatially dividing step is replaced with the prediction error signal obtained by the motion compensation predicting step performed immediately thereafter.

Similarly, if the prediction error signal is employed, a three-dimensional subband combining process in a moving picture decoding process

5 comprises:

the prediction error signal combining step of generating a combined subband prediction error signal by referring to a prediction error low-frequency signal which is a signal of a low-frequency band of the prediction error signal, and a high-frequency prediction error signal which is a subband 10 of high-frequency band adjacent to the low-frequency signal, and additionally both or either one of a low-frequency intrasubband which is in the same frequency band as the prediction error low-frequency signal, and a high-frequency intrasubband which is a subband of a high-frequency band adjacent to the low-frequency intrasubband, and the motion information 15 representing the motion compensation process corresponding to the prediction error signal;

the intraband signal spatially combining step of combining the low-frequency intrasubband and the high-frequency intrasubband; and

20 the motion compensation decoding step of performing a motion compensation predicting process on an intraband signal to add the combined prediction error signal thereto, thereby producing a decoded picture signal.

The prediction error signal combining step is performed on the prediction error low-frequency signal which is in the lowest frequency band of the prediction error signal, and the intraband signal spatially combining step 25 is performed on the low-frequency intrasubband which is in the lowest frequency band of the intraband signal. The band signal obtained by the

prediction error signal combining step is regarded as a new prediction error low-frequency signal, and the band signal obtained by the intraband signal spatially combining step is regarded as a new low-frequency intrasubband.

5 The prediction error signal combining step and the intraband signal spatially combining step are recursively repeated. As a result, the intraband signal and the prediction error signal are obtained.

The moving picture coding apparatus and the moving picture decoding apparatus described above can be implemented using a computer.

10 Specifically, the processing sequences and controlling processes of the moving picture coding apparatus and the moving picture decoding apparatus are realized when a program is executed by the computer. The computer mentioned here includes a processor and a controller. The program is read into the computer through a network or a recording medium such as a CD-ROM which stores the program. The present invention covers such a 15 program or a program product or a recording medium. A medium for transmitting such a program is also included in the scope of the present invention.

Fig. 8 shows an arrangement of a computer which implements the moving picture coding apparatus and/or the moving picture decoding apparatus. As shown in Fig. 8, the computer has processor 51, memory 52, and I/O interface 53 which are connected to each other by bus 54.

20 Memory 52 stores either one or both of a moving picture coding program and a moving picture decoding program which are to be executed by processor 51, and also serves as a temporary storage area while the 25 processor is executing the moving picture coding program or the moving picture decoding program. In this description, the term "memory" is used to

represent any of various memory devices such as a main memory unit such as a RAM, a cache memory included in a CPU or a register included in a processor, or a hard disk device. In the present embodiment, I/O interface 53 is a medium means for transmitting original pictures serving as an input to 5 and coded data serving as an output from the moving picture coding program under the control of processor 51, and also coded data serving as an input to and decoded pictures serving as an output from the moving picture decoding program under the control of processor 51. However, the presence of I/O interface 13 does not prevent the moving picture coding 10 method or the moving picture decoding method according to the present embodiment from being performed by storing original pictures and coded data, which are sought by another program, temporarily into memory 52 and reading them from memory 52.

Industrial applicability:

15 The present invention is applicable to uses wherein coded moving picture data are partially deleted from playback devices having various transmission environments and playback environments to allow moving picture distribution optimum for the environments of the playback devices.